03 | CAP 理论：分布式场景下我们真的只能三选二吗？

2022-01-26 陈现麟

《深入浅出分布式技术原理》 课程介绍 >

讲述：张浩

时长 14:35 大小 13.37M

你好，我是陈现麟。

通过学习分布式场景下面临的新挑战，你已经了解了从集中式系统演进到分布式系统时，分布式系统在部分故障、异步网络、时钟同步和共识协调这四个方面的新挑战，以及它们对分布式系统设计原则的影响。了解了这些之后，当你在面对分布式系统各种实现的时候，能更深刻地思考这些系统的取舍与权衡了。

经过不断地思考，人们在实践分布式系统架构的时候，从系统可用性和数据一致性的权衡中总结出来了 CAP 理论，它是指导人们在面对架构分布式系统时，进行取舍的设计原则。同时，CAP 理论深刻影响着分布式系统的设计与发展，是我们在学习分布式系统时不能绕过的知识。

所以在这节课中，我将和你一起来讨论什么是 CAP 理论以及它产生的影响，并且我们还会讨论在当前这个时间点，业界对于 CAP 理论的重新思考与理解。

什么是 CAP 理论

CAP 理论是加州理工大学伯克利分校的 Eric Brewer 教授在 2000 年 7 月的 ACM PODC 会议上首次提出的，它是 Eric Brewer 在 Inktomi 期间研发搜索引擎、分布式 Web 缓存时得出的关于数据一致性（ C：Consistency ）、服务可用性（ A：Availability ）、分区容错性（ P：Partition-tolerance ）的一个著名猜想：

> It is impossible for a web service to provide the three following guarantees : Consistency, Availability and Partition-tolerance.

在这个猜想提出的 2 年以后，来自麻省理工学院的 Seth Gilbert 和 Nancy Lynch 从理论上证明了 Eric Brewer 教授的 CAP 猜想是成立的，从此，CAP 理论在学术上正式成为了分布式领域公认的定理，并深刻影响着分布式系统的发展。

CAP 理论告诉我们，**一个分布式系统不可能同时满足数据一致性、服务可用性和分区容错性这三个基本需求，最多只能同时满足其中的两个**。为什么会这样呢？我们先来了解一下 CAP 理论对于数据一致性、服务可用性和分区容错性是怎么定义的。

一致性（ C ）

CAP 理论中的一致性是指强一致性（ Strong Consistency ），又叫线性一致性（ Linearizable Consistency ），它要求多节点组成的分布式系统，能像单节点一样运作，如果一个写操作返回成功，那么之后的读请求都必须读到这个新数据；如果返回失败，那么所有的读操作都不能读到这个数据。

一致性中除了强一致性之外，还有其他的一致性级别，比如序列一致性（ Sequential Consistency ）和最终一致性（ Eventual Consistency ）等，这个在后面的课程“一致性与共识（一）”中会有详细的介绍。

可用性（ A ）

CAP 理论对可用性的定义，指的是要求系统提供的服务必须处于 100% 可用的状态，对于用户的每一个操作请求，系统总能够在有限的时间内返回结果。**下面我们重点来讨论可用**

性定义中的三个关键点：“ 100% 可用”、“有限时间内”和“返回结果”。

第一点，100% 可用，既不是 99% 可用，也不是 99.99% 可用，它说的是系统必须完全可用，不允许任何不可用的情况出现，这是一个非常理想的模型。

第二点，有限时间内，它指的是对于客户端的一个请求，系统必须在指定的时间内返回对应的请求结果，如果超过了这个时间，系统就被认为是不可用的。一般来说，“有限时间内”是系统在设计的时候，就设定好的系统运行指标，不同的系统之间会有非常大的差别。

例如，对于一个服务在线业务的 OLTP 数据库 MySQL ，它的“有限时间”一般不会超过 1 秒，但是对于服务离线分析的 OLAP 数据库 Hive ，它的“有限时间”可能会超过 30 秒，甚至更长。

虽然不同的系统对于“有限时间”的设定差别非常大，但是对于一个给定的系统来说，在设定了这个“有限时间”之后，只要对一个请求的响应超过了这个时间，我们就认为这个系统是不可用的。

第三点，返回结果，这是指系统在完成对客户端请求的处理后，必须返回一个正常的响应结果。客户端可以根据这个响应结果，来明确判断这个请求执行成功还是失败，而不是返回一个让用户无法判断的不正常的响应结果。

比如客户发起一个请求，从用户 A 的账户转 50 元到用户 B 的账户，“转账成功”和“余额不足”都是正常的响应结果，而“服务不可达”和“服务器内部错误”等就是不正常的响应结果。

分区容错性（ P ）

分区指的是在整个分布式系统中，因为各种网络原因，系统被分隔成多个单独的部分，它不仅包含我们通常说的网络分区，也包含因为网络丢包导致的网络不通的情况。并且，这里说的因为网络丢包导致网络不通的情形，还包含节点宕机的场景，由于系统的其他机器不知道某个节点宕机了，只知道与宕机节点的网络是不通的，所以当节点宕机发生时，其他节点发往宕机节点的包也将丢失。

在现实的分布式系统中，我们面对的就是一个不可靠的网络和有一定概率宕机的设备，这两个因素都会导致分区出现，因此在分布式系统实现中，分区容错性 P 是一个必须项，而不是可选项。

在分布式系统中，如果我们的设计放弃分区容错性，就相当于我们认为节点之间的网络通信永远是好的，那么我们对节点之间的远程调用的结果，就不需要处理超时、网络地址不可达等网络层错误了。但是这样一来，看似是简化了系统设计，实际却忽视了超时等网络错误的情况。当它们出现后，系统的行为就是未定义的了，可能会出现崩溃，或者是脏数据的问题。

因此，对于分布式系统工程实践来说， CAP 理论更合适的描述是：在满足分区容错的前提下，没有算法能同时满足数据一致性和服务可用性。

CAP 理论产生的影响

关于数据一致性和可用性之间的争论由来已久，当时主要表现为 ACID 与 BASE 之间的争论。

基于 BASE 理论支撑的 NoSQL 运动坚持创造各种可用性优先、数据一致性其次的方案，而传统数据库则坚守 ACID 特性（原子性、一致性、隔离性、持久性），优先数据一致性，在必要的时候，可以放弃系统可用性。当时 BASE 理论还没有被广泛接受，人们还是不愿意放弃 ACID 的优点。

当 CAP 理论提出后，我们明白了在分布式系统中，只能在强一致性和 100% 的可用性之间二选一，不能两个都要。从此 BASE 理论也逐渐被人们所接受，在大规模存储的场景中广泛应用，并且开创了从 2000 年到 2010 年， NoSQL 运动的黄金十年。这十年里，工业界产生了大量优秀的 NoSQL 系统，比如 BigTable 、 HBase 、 MongoDB 、 Cassandra ，解决了人们当时遇到的大规模数据存储的问题。

CAP 理论的重新思考与理解

CAP 理论的出现是有历史使命的，让人们能够在分布式系统中，放弃以关系数据库为代表的 ACID 强一致性系统，接受以 NoSQL 为代表的 BASE 理论，并且暂时解决了人们在 2000 年前后对于分布式系统中，数据一致性和可用性之间的争论，让人们能够更加务实地解决当时由于互联网爆发式发展，产生的海量用户和数据的分布式计算与存储的问题。

一个有历史使命的事物，在使命完成后，要么就过时了，人们不再提起它，要么就会对它有新的解释，让它跟随时代一起发展下去。而 CAP 理论显然属于后者，因为直到现在，人们还在对它不断地重新思考与理解。

在 2000 年的时候，CAP 理论通过一个简单但是精确定义的模型，论证了在一个满足分区容错的分布式系统中，当我们进行系统设计时，只能在数据一致性和服务可用性之间二选一。其中，**数据一致性（ C ）指的是数据的强一致性，服务的可用性（ A ）指的是服务 100 % 的可用性，这才是 CAP 理论论证模型的关键点**。

对可用性的重新思考与理解

首先，我们对服务的可用性（ A ）进行分析，你会发现在我们的日常工作中，几乎没有见过 100% 可用的服务。可用性指标是在 0 到 100% 之间连续分布的，其实一个 100% 可用性的服务和一个 99.9999% 可用性的服务之间并没有多大的差别，如果我们的服务能实现 99.9999% 的可用性，哪怕它不符合 CAP 理论的可用性，也是符合我们工作中对可用性的要求的。

所以，在我们的系统选择了 CP 模型的时候，对于可用性（ A ），我们永远无法达到 100%，但是按业务要求不断优化，是我们努力的目标。

关于具体的实践，我认为基于 Raft 算法实现的 **etcd** 就是一个非常好的，对可用性进行重新思考的实践。如果依据 CAP 理论来划分的话，etcd 属于 CP 模型。

而在 etcd 系统的实现中，如果网络没有出现分区，整个系统是 100% 可用的；就算网络出现分区了，也不会有整个 etcd 系统都不可用的情况。在这时，超过半数 etcd 实例所在的网络分区一侧，系统是正常可用的，虽然网络分区的另一侧是不可用的，但是整个 etcd 系统的可用性依然可能超过 50% 。

对一致性的重新思考与理解

对于数据的一致性（ C ），除了 CAP 理论要求的强一致性外，还有单调一致性、会话一致性和最终一致性等。如果我们的系统设计选择了 AP 模型，**在数据一致性方面，虽然我们无法实现强一致性，但是我们也不要全部放弃，可以努力去实现更高的一致性级别，为系统的服务提供更好的抽象**。

这里我们通过一个例子来说明，假设我们设计一个 AP 模型的分布式系统，正常情况下，如果依据 CAP 理论，在系统设计时，我们需要放弃数据的一致性。但是，我们可以从另一个思路来设计，在系统没有出现网络分区的时候，这个分布式系统应该设计为强一致性的。

如果出现网络分区了，我们可以根据系统情况，有选择并且精心设计地降低系统的一致性级别。比如，从强一致性降低到单调一致性或会话一致性等，这样的设计，既符合 CAP 理论依据，也为系统提供了更好的一致性级别，特别是在网络分区的时候。

对分区容错性的重新思考与理解

最后，我们来分析一下分区容错性 P 的问题。在分布式系统中，节点之间必须通过网络来通信，可是网络可能会丢包和中断，节点也可能会宕机，这样的情况就要求我们在系统设计的时候，必须做好系统的分区容错处理。

但是，系统出现分区的情况非常少见，所以我们可以来试想一下，在网络不出现分区的时候，我们将数据强一致性和 100% 的可用性都选择，等到网络出现分区的时候，系统再选择放弃部分的可用性或者降低数据一致性的级别，这种处理方式是否可行呢？

其实这样的处理方式是可以的，在上面对可用性和一致性的重新思考与理解中，所举的例子都是按这个方式来处理的，**它实际是将 CAP 理论的选择，推迟到出现网络分区的时候，而不是系统一启动就进行 CAP 的选择**。这样可以大大提高系统的可用性和数据一致性，并且系统依然能容忍网络分区。

另外，关于 CAP 理论的重新思考，特别需要说明的一个例子是 Google 的 Spanner ，我们都知道 Spanner 是一个全球分布式数据库，但是 Google 却宣称 Spanner 是一个 CA 系统，这是不是和 CAP 理论的说法产生了矛盾呢？

其实并不矛盾，Spanner 虽然是一个分布式系统，但是它运行在 Google 的内部网络中，并且拥有大量冗余的网络链路、处理相关故障的架构规划、以及非常细致的运维，以此来确保系统的可用性超过了 99.999%。虽然不能达到 100%，但是对于使用者来说，和可用性 100% 几乎没有任何区别，所以 Spanner 就是一个 CA 系统。

而且，在网络出现分区的时候，Spanner 会选择一致性而不是可用性，这个时候 CAP 理论依然会生效。所以对于 CAP 理论的重新思考，总而言之就是一句话：**CAP 理论给我们定**

义了系统的设计边界，虽然想要设计出超过边界的系统是徒劳的，但是我们却可以无限逼近边界，并且把它作为我们设计系统的目标。

总结

到这里，我们知道了在分布式系统场景下，CAP 理论的相关知识，我们一起来总结一下这节课的主要内容：

首先，我们一起讨论了什么是 CAP 理论，它是指分布式系统中，在满足分区容错的前提下，没有算法能同时满足数据一致性和服务可用性，只能在数据一致性和服务可用性之间二选一。

然后，我们讨论了 CAP 理论产生的影响，可以说 CAP 理论的出现，让人们接受了 BASE 理论，并且推动了 NoSQL 运动的发展，开启了它的黄金十年。

最后，我们探讨了现在人们对于 CAP 理论的新理解，**对于 CAP 理论，我们不会简单地三选二或者二选一。**对于 AP 模型的系统，我们会努力去提升数据一致性的级别，而对于 CP 模型的系统，我们会努力去提升系统可用性的级别。

同时，由于系统分区的情况非常少见，我们可以在网络不出现分区的时候，将 A 和 C 都选择上；在网络出现分区的时候，再选择放弃部分的可用性，或者降低数据一致性的级别，通过推迟 CAP 选择来提高系统的可用性和数据一致性。

思考题

在分布式场景下，对于 CAP 理论，我们真的只能三选二吗？

欢迎你在留言区发表你的看法。如果这节课对你有帮助，也推荐你分享给更多的同事、朋友。

分享给需要的人，Ta订阅超级会员，你将得 50 元

Ta单独购买本课程，你将得 20 元

生成海报并分享

赞 3　提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇　02 | 新的挑战：分布式系统是银弹吗？我看未必！

下一篇　春节加餐 | 系统性思维，高效学习和工作的利器

精选留言 (5)

写留言

blentle

2022-01-27

cap从每一个定义来说并不是说牺牲了第三者吧，而是尽量保证第三者

作者回复: 是的，只是经常说成 3 选 2，让大家的感觉是要放弃 1 个

4

努力努力再努力

2022-01-26

1 一致性 （C）一致性级别 是 从高到低

1.1 强一致性

1.2 单调一致性

1.3 会话一致性

1.4 序列一致性...

展开

作者回复: 非常棒

4

GAC·DU

2022-01-27

分布式场景下，站在用户角度思考CAP理论，用户发送一条请求，在正确的时间响应正确的结果，可以被视为同时满足三者的系统设计。站在系统设计者的角度思考CAP理论，同时满足两者而第三者则是努力的方向，先让系统能够用起来，然后迭代优化。这也满足现在常说的"小步子快跑"的说法。

展开

作者回复: 非常棒！

2

三毛
2022-02-04

关于P的重新理解这块，没出问题的时候先上CA，也就是确保强一致性和服务高可用，当出问题的时候再改成CP或者AP，但是架构上不能同时满足吧？除非之前的CA也是尽量CA。

作者回复: 是的，文章中 Google 的 Spanner 就是一个比较好的例子

葡萄糖sugar
2022-01-27

作者你好，cp 我能理解为系统在达到强一致性以及网络分区容忍性，与此同时我们还能够尽力达到可用性，并且依旧保持强一致性吗？

作者回复: 以 CAP 的定义来说，A 是指 100% 的可用性，所以分布式系统中，CA 不可以兼得，不过可用性的范围在0%-100%之间，99.9% 或者更高的可用性都有可能实现的，比如专栏中提到的 spanner。

共 3 条评论 >